



- 1 -

TITLE OF THE INVENTION

SPEECH INPUT SYSTEM, SPEECH PORTAL SERVER, AND SPEECH INPUT
TERMINAL

5 DETAILED DESCRIPTION OF THE INVENTION]

[0001]

The present invention relates to a speech input system,
a speech portal server, and a speech input terminal, and more
specifically to a speech input system which have an access
10 from a mobile terminal device such as a portable phone, and
an onboard navigation system, and a home (stationary) terminal
such as a home telephone, a TV set, or a PC to an network with
a speech, and receive information and services from an
information service provider which provides map information,
15 music information, TV broadcast program information, and
telephone information.

[0002]

Japanese application patent laid-open publication No. Hei
11-143493 describes a system which converts a provided speech
20 into an intermediate language, which is a database language,
with a speech language understanding device, and searches a
word.

[0003]

Japanese application patent laid-open publication No.
25 2000-57490 describes a method for increasing a recognition
capability for a provided speech while switching recognition
dictionaries.

[0004]

Japanese application patent laid-open publication No. 2001-34292 describes a method for increasing a recognition capability where a word from a dictionary is extracted with a word spotting technique, a requested key word is recognized to determine a topic, and a speech is recognized with a recognition dictionary specific to the topic.

[0005]

The technique in Japanese application patent laid-open publication No. Hei 11-143493 is a method to learn a hidden Markov model which converts a sentence data into a corresponding intermediate language such that a recognition error becomes minimum. Since this method is a learning based on statistic processing, learning in individual fields is required when service is provided for different fields simultaneously, the processing takes a long time, and the recognition capability decreases. This is not designed as a speech input system considering an actual conversation which includes mixed long sentences and short sentences. Further, no attention is paid for a case where there is an error in a part of recognized string.

[0006]

The technique in Japanese application patent laid-open publication No. 2000-57490 is an invention for a navigation system for increasing a recognition capability while switching corresponding dictionaries according to a recognized result, and speech cannot be provided consecutively. No attention is paid for a case where there is an error in a part of recognized string.

[0007]

The technique in Japanese application patent laid-open publication No. 2001-34292 is an invention for increasing recognition capability while extracting a topic according to a recognized result, and switching dictionaries. No attention is paid for a case where there is an error in a part of recognized string as the two prior inventions described above.

SUMMARY OF THE INVENTION

10 [0008]

The purpose of the present invention is to provide a speech input system, a speech portal server and a speech input system which have access from a mobile terminal such as a PDA and a portable phone, and stationary terminal such as a home telephone (Home TEL), a TV set and a PC to a network with speech, and receives services from a provider for providing map information, music information, broadcast program information and telephone information.

[0009]

20 The present invention proposes a speech input system comprising a speech input terminal provided with a speech input/output mean, a Web browser, and a display mean for displaying an access status to an external system and a search result, a speech portal server provided with a speech
25 recognizing mean for receiving a speech from the speech input terminal to recognize it as a text, a command converting mean for checking the recognized text with a command text dictionary, and separating it into a command text and an object text, and

a conversation control mean for having an access to, and receiving a service from an application service provider which provides different information based on the separated command text and object text, and providing the speech input terminal with the service, and an application service provider which is provided with an information search mean for searching information based on the command text and the object text received from the speech portal server, and serves the speech portal server with a search result.

10 [0010]

The information search mean of the application service provider can extract every (n) characters from the received object text, and search for information based on an n-character INDEX created beforehand.

15 [0011]

The application service provider includes a navigation information application service provider for serving map information, a music information application service provider for serving music information, a broadcast program information application service provider for serving at least one type of information of TV broadcast program information, CS broadcast program information, and CATV broadcast program information, and a telephone information application service provider for serving telephone information.

25 [0012]

The speech portal server recognizes a speech received by the speech input terminal, and separates it into a command text and an object text, conducts a fuzzy search for information

stored in the application service provider based on the separated texts, and provides the speech input terminal with intended information even if there is a partial recognition error in the object text in the present invention.

5 [0013]

 The present invention propose a speech input system comprising a speech input terminal provided with a speech input/output mean, and a mean for displaying an access status to an external system, an application service provider for providing different information, and a speech portal server which controls a conversation between the speech input terminal and the application service provider based on the provided speech, and is provided with a speech recognizing mean for receiving a speech from the speech input terminal to recognize it as a text, a command converting mean for checking the recognized text with a command text dictionary, and separating it into a command text and an object text, and a conversation control mean for sending the separated command text and the object text to the application service provider, and providing the speech input terminal with information searched by said application service provider.

10
15
20

 [0014]

 The speech recognizing mean is provided with a connected speech recognizing mean, a word speech recognizing mean, and a comprehensive recognition evaluating mean for selecting either of recognition results of said two recognizing means with a speech characteristic value provided as a threshold.

25

 [0015]

The speech characteristic value is a speech time or a recognized string length.

[0016]

Since a speech recognizing engine in the speech portal server comprises a connected speech recognizing engine suitable for a long sentence, and a word speech recognizing engine suitable for a short sentence such as a command for comprehensive evaluation, thereby increasing recognition capability of speech conversation.

10 [0017]

A speech input terminal which has access to the speech portal server and the application service provider for providing different information is provided with a speech input/output mean, a Web browser, and a display mean for displaying an access status to an external system and a search result in the present invention.

[0018]

The speech input terminals are classified into portable speech input terminals which are integrated into any one of a PDA, a portable phone, or an onboard navigation system, and home speech input terminals which are integrated into any one of a home telephone, a TV set and a PC.

[0019]

Since the navigation information ASP, the music information ASP, the broadcast program information ASP, and the telephone information ASP are provided as application service providers (ASP's), the mobile speech input terminals such as a PDA, a Mobile TEL, and a Mobile Car PC and the home

speech input terminals such as a home telephone, a TV set, and a PC are served with optimal information

BRIEF DESCRIPTION OF THE DRAWINGS

5 FIG. 1 is a block diagram showing an overall constitution of an embodiment of a speech input system of the present invention.

FIG. 2 is a block diagram for showing an embodiment of a PDA serving as a speech input terminal of the present invention.

10 FIG. 3 is a block diagram for showing an embodiment of a Mobile TEL serving as a speech input terminal of the present invention.

FIG. 4 is a block diagram for showing an embodiment of a Mobile Car PC serving as a speech input terminal of the present invention.

15 FIG. 5 is a block diagram for showing an embodiment of a home telephone serving as a speech input terminal of the present invention.

FIG. 6 is a block diagram for showing an embodiment of a TV set serving as a speech input terminal of the present invention.

FIG. 7 is a block diagram for showing an embodiment of a PC serving as a speech input terminal of the present invention.

FIG. 8 is a block diagram for showing an embodiment of a speech portal server of the present invention.

25 FIG. 9 is a block diagram for showing a constitution of a speech recognizing mean of a speech portal server of the present invention.

FIG. 10 is a block diagram for showing an operation of a speech recognizing mean of a speech portal server of the present invention.

5 FIG. 11 is a block diagram for showing an operation of a speech recognizing mean of a speech portal server of the present invention.

FIG. 12 is a block diagram for showing an operation of a speech recognizing mean of a speech portal server of the present invention.

10 FIG. 13 is a block diagram for showing an operation of a speech recognizing mean of a speech portal server of the present invention.

15 FIG. 14 is a block diagram for showing a constitution of a command converting mean of a speech portal server of the present invention.

FIG. 15 is a drawing for showing an example of a speech command text dictionary of the present invention.

20 FIG. 16 is a drawing for showing an operation of a command converting mean of a speech portal server of the present invention.

FIG. 17 is a block diagram for showing a constitution of a conversation control mean of a speech portal server of the present invention.

25 FIG. 18 is a block diagram for showing a constitution of a navigation information ASP of the present invention.

FIG. 19 is a block diagram for showing a constitution of a music information ASP of the present invention.

FIG. 20 is a block diagram for showing a constitution of

a TV broadcast program information ASP of the present invention.

FIG. 21 is a block diagram for showing a constitution of a telephone information ASP of the present invention.

5 FIG. 22 is a drawing for showing an example of a speech operation menu screen of the present invention.

FIG. 23 is a block diagram for showing a constitution of a fuzzy search mean of individual information ASP's of the present invention.

10 FIG. 24 is a drawing for showing an example of a procedure for a fuzzy search mean of individual information ASP's of the present invention.

FIG. 25 is a drawing for showing a communication procedure among a speech input terminal, a speech portal server and a navigation information ASP of the present invention.

15 FIG. 26 is a drawing for showing a communication procedure among a speech input terminal, a speech portal server and a music information ASP of the present invention.

20 FIG. 27 is a drawing for showing a communication procedure among a speech input terminal, a speech portal server and a TV broadcast program information ASP of the present invention.

FIG. 28 is a drawing for showing a communication procedure among a speech input terminal, a speech portal server and a telephone information ASP of the present invention.

25 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0020]

The following section describes embodiments of a speech input system, a speech portal server, and a speech input terminal

while referring to FIG. 1 to FIG. 28.

[0021]

FIG. 1 is a block diagram showing an overall constitution
of an embodiment of a speech input system of the present
5 invention.

[0022]

A mobile terminal 10 and a home (stationary) terminal 30
are available as a speech input terminal device in the present
embodiment. The mobile terminal 10 includes a PDA 10a, a
10 cellular phone 10b, and an onboard terminal 10c. The home
(stationary) terminal 30 includes a stationary telephone 30a
which is intended for a household application, a television
system TV 30b as an information home electric appliance, and
a personal computer PC 30c. The portable terminals 10a to 10c
15 are connected with an Internet network 40 through a radio base
station 20, and the home terminals 30a to 30c are directly
connected with the Internet network 40.

[0023]

The speech portal 50 for controlling the entire speech
20 conversation, and different application service providers
(ASP's) 60 are connected with the Internet network 40 as well.

[0024]

The ASP includes a navigation information ASP 60a for
serving map information, a music information ASP 60b, a TV
25 broadcast program information ASP 60c, and a telephone
information ASP 60d.

[0025]

When any one of the speech input terminals 10a to 10c,

and 30a to 30c connects with a speech portal server 50, a speech guidance and a menu display are provided on the speech input terminal, and entering a corresponding speech transmits the speech to the speech portal server 50 through the Internet
5 network 40.

[0026]

The speech portal server 50 recognizes the speech, applies the command conversion to the content of the speech for converting it into a command and an object to be searched for,
10 and transmits them to an ASP 60 corresponding to the content of the command.

[0027]

The ASP 60 searches a corresponding database, and provides the speech input terminal where the speech is entered with
15 a search result through the speech portal server 50.

[0028]

As describe above, the speech input system is mainly used for a potable terminal under an environment where a keyboard (KB) is hardly available, and a household where keyboard
20 operation is not popular, thereby facilitating the input.

[0029]

A group of servers are connected with the Internet as an overall constitution of the speech input system in the present embodiment.

25 [0030]

When the group of servers are connected with an intranet or a home network, their effect does not make a difference in these networks. It is possible that different types of ASP

group are provided closely, and the ASP group serves as a cache server which is connected with the Internet server group only when the ASP group cannot provide an intended service.

[0031]

5 Information services other than those shown in FIG. 1 such as stock price information, trading partner information, customer information and product information may exist in the ASP server group.

[0032]

10 The speech portal server 50 may manage personal information, and provide services according to personal characteristics.

[0033]

15 FIG. 2 to FIG. 4 show constitutions of mobile terminals, and FIG. 5 to FIG. 7 show constitutions of home (stationary) terminals. A primary part of the individual terminals is constituted almost in the same way.

[0034]

20 FIG. 2 is a block diagram for showing an embodiment of a PDA serving as a speech input terminal of the present invention. The PDA 10a includes an antenna for communicating with the radio base station 20, and a communication mean 10a2 for the radio communication. The communication mean 10a2 can transmit and receive speeches and data simultaneously with a Voice Over IP (VoIP) technology or the like. A processing apparatus and
25 a Web browser 10a3 are connected with individual constituting parts and peripherals, and control the entire terminal. The peripherals include a microphone MIC 10a4 for speech input, a coordinate input apparatus (tablet) TB 10a5 constituted as

a touch panel, a liquid crystal display LCD 10a6, and a speaker SP 10a7.

[0035]

5 The PDA 10a is provided with a position detecting mean 10a8 important for a mobile terminal, and is connected with the GPS (Global Positioning System) 10a9.

[0036]

10 The touch panel and a speech can be used for the operation of the PDA 10a. The processed result is shown on the display, and the PDA 10a enters a state ready for the next operation.

[0037]

15 FIG. 3 is a block diagram for showing a constitution of a Mobile TEL serving as a speech input terminal of the present invention. The constituting elements are the same as those for the PDA 10a in FIG. 2. The size and the color display capability of a liquid crystal display LCD 10b are generally different for reducing its cost. On the other hand, different types of software for a mobile telephone is added.

[0038]

20 FIG. 4 is a block diagram for showing a constitution of a Mobile Car PC serving as a speech input terminal of the present invention. The constituting elements are basically the same as those for the PDA 10a in FIG. 2. A liquid crystal display LCD 10c6 suitable for installing on a vehicle and onboard application software are different from the PDA 10a in FIG. 25 2. The Mobile Car PC is connected with different onboard sensors, which are not shown in the figure, and may show information on a vehicle.

[0039]

FIG. 5 is a block diagram for showing a constitution of a home telephone serving as a speech input terminal of the present invention. The difference from the PDA 10a shown in
5 FIG. 2 is that it does not have the antenna 10a1 for communicating with the radio base station 20, the position detecting mean 10a8, and the GPS 10a9.

[0040]

FIG. 6 is a block diagram for showing a constitution of
10 a TV set serving as a speech input terminal of the present invention. The difference from the stationary home telephone 30a in FIG. 5 includes a television apparatus TV 30b10, a TV set control mean 30b8, and a camera CM 30c9. The TV control mean 30b8 is a mean for programming for recoding a TV broadcast
15 program, and setting a channel, and is generally referred as a set top box.

[0041]

The camera CM 30c9 is used for transmitting an image for conversation to the other party, and monitoring a room with
20 an image.

[0042]

FIG. 7 is a block diagram for showing a constitution of a PC serving as a speech input terminal of the present invention. The difference from the stationary home telephone 30a in FIG.
25 5 is that there is no TV set control mean. The operation for the PC is conducted through a touch panel or a speech. A keyboard suppressed from the drawing may be connected for operating the PC.

[0043]

The camera CM 30c9 shown in FIG. 6 and FIG. 7 may be installed on the speech input terminal in FIG. 2 to FIG. 5.

[0044]

5 FIG. 8 is a block diagram for showing a constitution of
an embodiment of a speech portal server 50 of the present
invention. The speech portal server 50, which is a
characteristic part of the present invention, comprises a
communication mean 501 for communicating with the Internet
10 network 40, a processing apparatus 502 for processing the entire
speech portal server 50, a speech recognizing mean 503 for
receiving speech data Vin, recognizing it with a recognition
dictionary 504, and providing text data Vtext1, a command
converting mean 505 for using a command text dictionary 506
15 to convert the recognized speech Vtext1 into a command and
an object Vtext2, a conversation control mean 507 for
controlling a conversation with the speech input terminal and
different information ASP's, a speech synthesizing mean 508
for synthesizing a speech with a speech text Vtext3 from the
20 conversation control mean, and a Web browser 509.

[0045]

FIG. 9 is a block diagram for showing a constitution of
the speech recognizing mean 503 of the speech portal server
50 of the present invention. The present embodiment features
25 that the speech recognizing mean is provided with two
recognizing engines. Namely, the speech recognizing mean 503
comprises a connected speech recognizing engine 503a for
recognizing a relatively long speech, and a word speech

recognizing engine 503b for recognizing a relatively short speech such as a command.

[0046]

5 The connected speech recognizing engine 503a uses a connected speech recognition dictionary 504a for recognizing a speech, and the work speech recognizing engine 503b uses a word speech recognition dictionary 504b for recognizing a speech.

[0047]

10 A comprehensive recognition evaluating mean 503c comprehensively evaluates recognition results of the individual recognizing engines. Generally, the connected speech recognizing engine uses a recognizing method for using a transition probability model between words, and presents
15 an increased number of recognition errors when a short word such as a command is entered since knowledge on the preceding and the following words is not used.

[0048]

20 Thus, it is required for the comprehensive recognition evaluating mean to comprehensively determine which output from the recognizing engines is correct.

[0049]

The following section specifically describes the operation of the comprehensive recognition evaluating mean 503c while
25 using a specific example, and referring to FIG. 10 to FIG. 13

[0050]

FIG. 10 to FIG. 11 show examples where a speech time is

used to switch the recognition results from the two recognizing engines. The comprehensive recognition evaluating mean 503c compares the speech data Vin with a threshold, and switches to S when Vin short and switches to L when Vin is long in the
5 speech time evaluation.

[0051]

FIG. 10 is a figure for describing an operation of the speech recognizing mean of the speech portal server 50 of the present invention, and shows a state where a speech "はい。" with
10 a relatively short speech time is entered. In this case, the comprehensive recognition evaluating mean is switched to the S side, and provides an output Vtext1 of a string "はい。". Here, the maximum speech time from the word speech recognition dictionary is selected for the threshold.

15 [0052]

FIG. 11 is a figure for describing an operation of the speech recognizing mean of the speech portal server 50 of the present invention, and shows a state where a speech data "日立神田に目的地を設定する。" with a relatively long speech time is
20 entered. In this case, the comprehensive recognition evaluating mean is switched to the L side, and provides an output Vtext1 of a string "日立神田に目的地を設定する。".

[0053]

FIG. 12 to FIG. 13 show examples where a speech time is
25 not evaluated, but the resultant string length from the recognizing engine is compared with a threshold.

[0054]

FIG. 12 is a figure for describing an operation of the

speech recognizing mean of the speech portal server 50 of the present invention, and shows a state where a speech "はい。" with a relatively short speech time is entered. In this case, the comprehensive recognition evaluating mean is switched to the S side, and provides an output Vtext1 of a string "はい。".

[0055]

FIG. 13 is a figure for describing an operation of the speech recognizing mean of the speech portal server 50 of the present invention, and shows a state where a speech data "日立神田に目的地を設定する。" with a relatively long speech time is entered. In this case, the comprehensive recognition evaluating mean is switched to the L side, and provides an output Vtext1 of a string "日立神田に目的地を設定する。".

[0056]

It is set that the engine provides a string "認識不能を示す文字列="?" for indicating that the recognition is impossible when the recognizing engine receives a speech which is largely different from the dictionary. In this case, when the threshold is selected to a proper value (example: the maximum length of the command strings), an optimal string is provided, thereby improving the overall recognition capability.

[0057]

As describe in the two types of methods above, when the speech data is a command "はい。", a problem that a string "肺。" is provided when there is only the connected speech recognizing engine is solved.

[0058]

FIG. 14 is a block diagram showing a constitution of the

command converting mean 505 of the speech portal server 50
of the present invention. When the command converting mean
receives a string Vtext1 from the speech recognizing mean 503,
a command string search 505a uses the command text dictionary
5 to determine if a command string is included.

[0059]

FIG. 15 is a figure for showing an example of the speech
command text dictionary of the present invention. The command
text dictionary 506 includes a command ID and command name
10 1 to command name 5, and any commands with the same command
ID are treated in the same way. A string "目的地設定。" and a
string "行く。" are determined as an identical command ID, D01.

[0060]

Command text strings in FIG. 15 are roughly classified
15 into commands corresponding to the individual information ASP's
(NO 1 to NO 8), commands for speech conversation (NO 9 to NO
10), and commands for screen operations (NO 11 to NO 22).

[0061]

Though command string search assumes a complete match,
20 it may be designed that the search can be conducted for a case
where a partial error is included as described later in FIG.
23 to FIG. 24.

[0062]

Object extraction 505b for extracting a string of an object
25 other than a command is conducted after the command string
search 505a. This processing is a processing for extracting
a command transmitted to the individual APS's, and an object
which is a subject of the search.

[0063]

FIG. 16 is a drawing for describing an operation of the command converting mean 505 of the speech portal server 50 of the present invention. When a result Vtext1 of the speech
5 recognizing mean is "日立神田に目的地を設定する。", the command text search 505a refers to the command text dictionary 506, determines that the command string is "目的地を設定する。", and determines that the command ID is D01.

[0064]

10 Then the object extraction 505b determines that the part other than the command string is an object, and an object "日立神田に。" is extracted. Thus, a result Vtext2 of the object extracting mean 505b is provided as "command ID=D01, Object=日立神田に。".

15 [0065]

Though the entire string other than the command characters is assumed as an object in the object extracting mean 505b, it is possible to conduct a morpheme analysis to remove "に" from "日立神田に。" for the extraction.

20 [0066]

FIG. 17 is a block diagram for showing a constitution of a conversation control mean 507 of the speech portal server 50. The conversation control mean 507 comprises a conversation processing mean 507a for controlling the entire part, basic
25 conversation rules 507b for speech conversation, a terminal data control mean 507c for serving as an interface with the speech input terminal, an ASP control mean 507d for serving as an interface with the individual information ASP's, and

a speech synthesizing control mean 507e.

[0067]

The basic conversation rules 507b store rules commonly used among the individual information ASP's, and conversation
5 rules specific to the individual information ASP's are downloaded from the individual information ASP's.

[0068]

When the conversation processing mean 507a receives Vtext2 of output from the speech recognizing mean 505, determines
10 a command ID, determines which information ASP it corresponds to, and transmits the command ID and an object as ASPDataOut to the corresponding information ASP.

[0069]

When the corresponding ASP provides ASP control mean 507b
15 with a search result as ASPDataIn, the terminal control 507c provides the speech input terminal which has requested the search with TdataOut. The speech input terminal shown the data of the search result.

[0070]

20 When a synthesized speech from the string is provided, the speech synthesizing control mean 507e provides a speech sequence as Vtext3, a speech Vout synthesized by the speech synthesizing mean 508 is transmitted to the speech input terminal, and a speaker provides sounds.

25 [0071]

When there is a data input other than a speech from the speech input terminal, it is received as TdataIn.

[0072]

The conversation control mean 507 may be constituted with a Voice XML browser for speech conversation.

[0073]

The following section describes constitutions of the individual information ASP's while referring to FIG. 18 to FIG. 21.

[0074]

FIG. 18 is a block diagram for showing a constitution of a navigation information ASP of the present invention. The navigation information ASP is a provider serving map information and path search information, and comprises an interface 60a100, a fuzzy search mean 60a200, a path search mean 60a500, and a conversation rule processing mean 60a700. The individual means refer to individual dictionaries to process requests.

[0075]

The fuzzy search mean 60a200 refers to a landmark DB 60a300, which is a database for landmark information, and a landmark INDEX 60a400 dictionary for the fuzzy search. The detailed operation is described later.

[0076]

The path search mean 60a500 refers to Map DB 60a600 which are map data, and searches a path from a current position to a destination. This path search is a general path search processing, and the detailed description is skipped.

[0077]

The conversation rule processing mean 60a700 is a mean for processing conversation rules specific to the individual

information ASP's, and the conversation rules 60b800 are used as rules other than base conversation rules for the speech portal server 50.

[0078]

5 FIG. 19 is a block diagram for showing a constitution of a music information ASP of the present invention. Compared with the ASP in FIG. 18, the music information ASP does not include what corresponds to the path search mean, and its contents are a music DB 60b300, a music INDEX 60b400, and
10 conversation rules for music 60b800.

[0079]

FIG. 20 is a block diagram showing a constitution of a TV broadcast program information ASP of the present invention. It differs from the ASP in FIG. 19 in contents. Its contents
15 include a TV broadcast program DB 60c300, a broadcast program INDEX 60c300, a broadcast program INDEX 60c400 and rules for broadcast programs 60c800.

[0080]

The TV broadcast programs mean at least one type of
20 information of information on TV broadcast programs, information on CS broadcast programs, and information of CATV broadcast programs in this specification.

[0081]

FIG. 21 is a block diagram for showing a constitution of
25 a telephone information ASP of the present invention. It has contents different from those for the ASP in FIG. 19, and is provided with a telephone DB 60d300, a telephone INDEX 60d400, and conversation rules for telephone 60d800.

[0082]

FIG.22 shows an example for a speech operation menu screen of the present invention. The speech menu is provided with speech menu icons according to the individual information ASP's.

5 The icons relating to the navigation information ASP includes "目的地設定。", "目的地検索。", "経由地設定。", and "現在地表示。".

[0083]

10 The icons relating to the music information ASP includes "音楽検索.". The icons relating to the broadcast program ASP includes "番組検索。", "番組予約.". The icons relating to the telephone information ASP includes "電話検索。".

[0084]

15 The present invention allows a method for entering an item from the speech menu and a method for entering everything including an object. For example, it is possible to enter a speech of "日立神田に目的地を設定する." for a destination search without pressing the menu.

[0085]

20 FIG. 23 is a block diagram for showing a constitution of the fuzzy search mean 60a200 for the individual information ASP's of the present invention. The other fuzzy search means 60b200, 60c200, and 60d200 have a constitution same as that of the fuzzy search mean 60a 200.

25 [0086]

The fuzzy search mean 60a200 comprises a search engine 60a100 and two-character INDEX generation 60a220 in FIG. 23. The search engine 60a100 and the two-character INDEX generation

60a220 search while referring to the landmark DB 60a300, and the landmark INDEX 60a400.

[0087]

Since the landmark DB stores large amount of data up to
5 several million of items, it is required that the landmark
DB has generated the two-character INDEX before hand. The
present invention features a fast and fuzzy search with this
two-character INDEX generation processing. The fuzzy search
here does not mean that meaning is fuzzy but that entered words
10 are searched when there is a partial error in a string (a
partially added string, a partially missing string, a random
string order, and a partial error string).

[0088]

FIG. 24 shows an example of the fuzzy search for the
15 individual information ASP's of the present invention. This
example sets a destination to "日立神田に。".

[0089]

First , a search key word 60a300key of "日立神田に。" is
entered, a processing for extracting every two characters
20 60a211 is conducted.

[0090]

Then, a landmark INDEX search 60a212 is conducted for the
every two characters.

[0091]

25 The landmark INDEX search 20a212 searches the landmark
DB, and extracts hit DB records.

[0092]

The extracted records are sorted in the order of the number

of hit characters, and an output processing 60a214 is conducted,
and a list 60a200res is provided as a search result.

[0093]

Since the INDEX search by every two characters is conducted
5 as described above, it has characteristics of the fast search
and the fuzzy search.

[0094]

When "に。" of "日立神田に。" is included in a search object,
if the landmark DB does not has a correspondence, it ignores
10 it. On the other hand, even if "日立神田に。" is entered, what
relating to it hit.

[0095]

Thus, it has an effect that a name of a place or a landmark
which occurs to mind can be entered.

15 [0096]

It is also possible to design such that multiple search
results are shown on the speech input terminal, and a speech
instruction prompts for selecting the search subjects.

[0097]

20 Though the present embodiment uses two-character INDEX
generation processing for the search as shown in FIG. 23 to
FIG. 24, a three-character INDEX, or a four-character INDEX
is also applicable.

[0098]

25 When information includes many numbers and alphabets, the
three-character INDEX generation or the four-character INDEX
generation presents less unnecessary search output compared
with the two-character INDEX processing.

[0099]

The following section describes specific communication procedures among the speech input terminal, the speech portal server 50, and the information ASP 60 while referring to FIG.

5 25 to FIG. 28.

[0100]

FIG. 25 shows a communication procedure among the speech input terminal, the speech portal server, and the navigation information ASP of the present invention. Here, the

10 communication procedure among the speech input terminal Mobile PC 10c, the speech portal server 50, and the navigation information ASP 60a, and a communication procedure with another information ASP is almost similar.

[0101]

15 First, when the speech input terminal Mobile PC 10c sends a connection request to the speech portal server 50, the speech portal server 50 provides the speech input terminal Mobile PC 10c with a speech output of "ご用件を入力ください。" through speech. Simultaneously, the speech menu in FIG. 21 is shown.

20 [0102]

Then, the speech input terminal Mobile PC 10c conducts a direct speech input of "日立神田に目的地を設定する。" through speech.

[0103]

25 The speech portal server 50 recognizes it, and responds to it as "日立神田に目的地を設定しますか?" through speech.

[0104]

Then, a speech for a command of "はい。" or "いいえ。" is

entered.

[0105]

When "はい。" is entered, the speech portal server 50 returns a speech response of "検索中です。" to the speech input terminal
5 Mobile PC 10c, sends data comprising a command ID "D01" and an object "日立神田に。" to the navigation information ASP 60a, and receives a search result. Here, a search result (two hits), and a content (XXX, YYY) are returned.

[0106]

10 The speech portal server 50 responds through speech as "結果が2件あります。何番にしますか?" according to the search result. Simultaneously the display of the speech input terminal Mobile PC 10c shows the content of the search result.

[0107]

15 Then, a speech instruction "1番。" is entered through speech, the speech portal server 50 recognizes the speech, and provides a corresponding speech output of "目的地を1番に設定します。".

[0108]

Further, it request the speech input terminal Mobile PC
20 10c for a current position, obtains current position information, and sends a path search command and its parameters to the navigation information ASP 60a based on this information.

[0109]

The speech portal server 50 receives path information and
25 map information as the search result from the navigation ASP, and provides the speech input terminal Mobile PC 10c with it, and responds as "安全運転でどうぞ。" through speech.

[0110]

When "いいえ。" is entered in the communication procedure described above, the procedure returns to "ご用件を入力ください。" again, which is suppressed from the drawing.

[0111]

5 When a selection is made from multiple search results, though an example for selecting through speech is presented, a touch panel is provided, and the touch panel is used for the selection. If this is the case, a correspondence between the content of the search and the coordinate of the touch panel
10 must be determined before hand.

[0112]

 Though the Mobile PC 10c is used as the speech input terminal, the PDA 10a, and the mobile TEL 10b can communicate with the navigation information ASP in FIG. 25. In this case, the
15 navigation system is a human. Since the current position of the own speech input terminal is known, it is possible to shown the current position information and search for a destination landmark.

[0113]

20 FIG. 26 is a drawing for showing a communication procedure among the speech input terminal, the speech portal server, and the music information ASP. This is a communication procedure where a speech input Mobile PC 10c receives a music content service from the music information ASP 60b through
25 the speech portal server 50.

[0114]

 First, when the speech input terminal Mobile PC 10c sends a connection request to the speech portal server 50, the speech

portal server 50 provides the speech input terminal Mobile PC 10c with a speech output of "ご用件を入力ください。" through speech. Simultaneously, the speech menu in FIG. 21 is shown.

[0115]

5 Then, the speech input terminal Mobile PC 10c conducts a direct speech input of "Mariah Careyの曲を聴きたい。" through speech.

[0116]

10 The speech portal server 50 recognizes it, and responds to it as "Mariah Careyの曲を検索しますか?" through speech.

[0117]

 Then, a speech for a command of "はい。" or "いいえ。" is entered.

[0118]

15 When "はい。" is entered, the speech portal server 50 returns a speech response of "検索中です。" to the speech input terminal Mobile PC 10c, sends data comprising a command ID "M01" and an object "Mariah Careyの。" to the music information ASP 60b, and receives a search result.

20 [0119]

 Here, a search result (three hits), and a content (XXX, YYY, ZZZ) are returned.

[0120]

25 The speech portal server 50 responds through speech as "結果が3件あります。何番にしますか?" according to the search result. Simultaneously the display of the speech input terminal Mobile PC 10c shows the content of the search result.

[0121]

Then, a speech instruction "3番." is entered through speech, the speech portal server 50 recognizes the speech, and provides a corresponding speech output of "音楽を3番に設定します".

Simultaneously, the music information ASP is instructed to
5 download the third music tune.

[0122]

With this, a speech response of "お楽しみください。" and a corresponding music content are downloaded to the speech input terminal Mobile PC 10c.

10 [0123]

When a search result includes only one item, after receiving a response of whether OK or not, a download starts.

[0124]

FIG. 27 shows a communication procedure among the speech
15 input terminal, the speech portal server, and the TV broadcast program information ASP of the present invention. This is a communication procedure where a speech input terminal TV 30b and a PC 30C receive a TV broadcast program content service from the broadcast program information ASP 60c through the
20 speech portal server 50.

[0125]

First, when the speech input terminal sends a connection request to the speech portal server 50, the speech portal server 50 provides the speech input terminal with a speech output
25 of "ご用件を入力ください。" through speech. Simultaneously, the speech menu in FIG. 21 is shown.

[0126]

Then, the speech input terminal conducts a direct speech

input of "天気予報の番組を見たい。" through speech.

[0127]

The speech portal server 50 recognizes it, and responds to it as "天気予報の番組を検索しますか?" through speech.

5 [0128]

Then, a speech for a command of "はい。" or "いいえ。" is entered.

[0129]

When "はい。" is entered, the speech portal server 50 returns
10 a speech response of "検索中です。" to the speech input terminal,
sends data comprising a command ID "T01" and an object "天
気予報の。" to the broadcast program information ASP 60c, and
receives a search result.

[0130]

15 Here, a search result (two hits), and a content (XXX, YYY)
are returned.

[0131]

The speech portal server 50 responds through speech as "結
果が2件あります。何番にしますか?" according to the search result.
20 Simultaneously the display of the speech input terminal shows
the content of the search result.

[0132]

Then, an speech instruction "1番。" is entered through speech,
the speech portal server 50 recognizes the speech, and provides
25 a corresponding speech output of "番組を1番に設定します。".

[0133]

As the result, a channel corresponding to the TV broadcast
program is set, thereby allowing viewing a weather forecast

service.

[0134]

When a search result includes only one item, after receiving a response of whether OK or not, a channel is set.

5 [0135]

When a weather forecast service is not being broadcasted, a channel can be programmed. In this case, the speech portal server 50 provides a guidance asking if programming is conducted or not, and the programming ends when it is responded.

10 [0136]

For TV broadcast programs viewed every week, programming for every week is available.

[0137]

FIG. 28 shows a communication procedure among the speech
15 input terminal, the speech portal server, and the telephone
information ASP of the present invention. This is a
communication procedure where a speech input terminal home
telephone 30a and a Mobile TEL 10b receive a telephone
information content service from the telephone information
20 ASP 60d through the speech portal server 50.

[0138]

First, when the speech input terminal sends a connection
request to the speech portal server 50, the speech portal server
50 provides the speech input terminal with a speech output
25 of "ご用件を入力ください。" through speech. Simultaneously, the
speech menu in FIG. 21 is shown.

[0139]

Then, the speech input terminal conducts a direct speech

input of "日立太郎さんに電話したい。" through speech.

[0140]

The speech portal server 50 recognizes it, and responds to it as "日立太郎さん電話をかけますか?" through speech.

5 [0141]

Then, a speech for a command of "はい。" or "いいえ。" is entered.

[0142]

When "はい。" is entered, the speech portal server 50 returns
10 a speech response of "検索中です。" to the speech input terminal,
sends data comprising a command ID "P01" and an object "日
立太郎さんに。" to the telephone information ASP 60d, and receives
a search result.

[0143]

15 Here, a search result (two hits), and a content (XXX, YYY)
are returned.

[0144]

The speech portal server 50 responds through speech as "結
果が2件あります。何番にしますか?" according to the search result.
20 Simultaneously the display of the speech input terminal shows
the content of the search result.

[0145]

Then, a speech instruction "1番。" is entered through speech,
the speech portal server 50 recognizes the speech, and provides
25 a corresponding speech output of "電話を1番にかけます。".

[0146]

Specifically, when a search result includes only one item,
after receiving a response of whether OK or not, a telephone

call is conducted.

[0147]

With the present invention, a speech portal server recognizes a speech entered from a speech input terminal, the
5 speech is separated into a command text and an object text, information stored in an application service provider is searched with fuzziness based on the separated texts, and an intended information is provided for the speech input terminal even if there is a partial error in the object text.

10 [0148]

Also, since a speech recognizing engine of the speech portal server is constituted such that it comprises two speech
recognizing engines of a connected speech recognizing engine
suitable for a long sentence, and a word speech recognizing
15 engine for a short sentence such as a command for comprehensive evaluation, the recognition capability for speech conversation increases.

[0149]

Further, since a navigation information ASP, a music
20 information ASP, a broadcast program information ASP, and a telephone information ASP are provided as application service providers (ASP's), mobile speech input terminals such as a PDA, a Mobile TEL, and a Mobile Car PC, and home speech input terminals such as a home telephone, a TV set, and a PC can
25 receive optimal information according to their individual requirement.

WHAT IS CLAIMED IS;

1. A speech input system comprising:

a speech input terminal provided with a speech input/output mean, a Web browser, and a display mean for displaying an access status to an external system and a search result;

a speech portal server provided with a speech recognizing mean for receiving a speech from said speech input terminal to recognize it as a text, a command converting mean for checking the recognized text with a command text dictionary, and separating it into a command text and an object text, and a conversation control mean; and

the application service provider which is provided with an information search mean for searching information based on the command text and the object text received from said speech portal server, and serves said speech portal server with a search result.

wherein said conversation control mean has access to, and receives a service from said application service provider which provides different information based on the separated command text and object text, and provides said speech input terminal with the service.

2. The speech input system according to Claim 1 wherein said information search mean of the application service provider extracts every (n) characters from the received object text, and searches for information based on n-character INDEX created beforehand.

3. The speech input system according to Claim 1 or Claim 2 wherein said application service provider is a

navigation information application service provider which serves map information.

4. The speech input system according to any one of Claim 1 to Claim 3 wherein said application service provider
5 is a music information application service provider which serves music information.

5. The speech input system according to any one of Claim 1 to Claim 4 wherein said application service provider is a broadcast program information application service provider
10 which serves information on at least one of a TV broadcast program, a CS broadcast program, and a CATV broadcast program.

6. The speech input system according to any one of Claim 1 to Claim 5 wherein said application service provider is a telephone information application service provider which
15 serves telephone information.

7. A speech input system comprising:

a speech input terminal provided with a speech input/output mean, and a mean for displaying an access status to an external system;

20 an application service provider for providing different information; and

a speech portal server which controls a conversation between said speech input terminal and said application service provider based on the provided speech, and is provided with
25 a speech recognizing mean for receiving a speech from said speech input terminal to recognize it as a text, a command converting mean for checking the recognized text with a command text dictionary, and separating it into a command text and

an object text, and a conversation control mean for sending the separated command text and object text to said application service provider, and providing said speech input terminal with information searched by said application service provider.

5 8. The speech portal server according to Claim 7 wherein said speech recognizing mean is provided with a connected speech recognizing mean, a word speech recognizing mean, and a comprehensive recognition evaluating mean for selecting either of recognition results of said two recognizing
10 means with a speech characteristic value provided as a threshold.

 9. The speech portal server according to Claim 8 wherein said speech characteristic value is a speech time.

 10. The speech portal server according to Claim 8
15 wherein said speech characteristic value is a recognized string length.

 11. A speech input terminal having an access to and receiving a service from a speech portal server and an application service provider for providing different
20 information, and provided with an speech input mean, a Web browser, and a display mean for displaying an access status to an external system and a search result.

 12. The portable speech input terminal according to Claim 11 integrated into any one of a PDA, a portable phone,
25 and an onboard navigation system.

 13. A home speech input terminal according to Claim 11 integrated into any one of a home telephone, a TV set, and a PC.

ABSTRACT OF THE DISCLOSURE

In order to provide a speech input system for having access from a mobile terminal such as a PDA or a portable phone, or a stationary terminal such as a home telephone, a TV set, or
5 a PC to a network through speech, and receiving a service from a provider for providing map information, music information, broadcast program information, and telephone information, the speech input system comprises speech input terminals 10, 30 provided with a speech input/output mean, and an access status
10 display mean, a speech portal server 50 provided with a speech recognizing mean for receiving a speech to recognize it as a text, a command converting mean for checking the recognized text with a command text dictionary, and separating it into a command text and an object text, and a conversation control
15 mean for having an access to, and receiving a service from a provider which provides different information based on the separated texts, and providing the speech input terminal with the service, and a provider 60 for searching information based on the command text and the object text received from the speech
20 portal server, and serves the speech portal server with a search result.